

Capstone Project Report

Battle of Neighborhoods

Introduction:

The city of **Los Angeles** (also known simply as L.A., and nicknamed the "City of Angels") is the most populous city in California. Located on a broad basin in Southern California, the city is surrounded by vast mountain ranges, valleys, forests, beautiful beaches along the Pacific Ocean, and nearby desert.

The metropolitan area is the second-most populous in the United States and home to over 17 million people who hail from all parts of the globe. The metropolitan area is spread across Los Angeles County, Orange County, and parts of San Bernardino County, Riverside County, and Ventura County.

Los Angeles is an important center of culture, medicine, agriculture, business, finance, energy, aerospace, science, food processing, media, international trade, and tourism. International tourists regard Los Angeles as most famous for "Hollywood".

The people of Los Angeles come from all over the world and are dispersed throughout the city's many sprawling, unique neighborhoods, though many of them congregate in ethnic enclaves like Little Armenia, Koreatown, Little Ethiopia, Chinatown, Little Tokyo, Historic Filipinotown or Tehrangeles.

Problem statement:

Many people visit city of Los Angeles daily. Many explorers and tourists have constraint over the time and money they can invest in their travels. But any visitor to city would like to experience the different folds of the city without having repetitive and same experiences.

Hence it would be ideal for people if they knew the top locations to visit the in places across the city and having to know the cluster of similar neighborhood and those which are different to other, which they have not visited.

Data description:

The required data took lot of digging over various websites and forms.

To gain the required data we used and scrapped the data from primarily two sites.

1) http://www.laalmanac.com/communications/cm02_communities.php

2) <https://gist.githubusercontent.com/senning/58a8c82e0c97712eabbe4700ce2187a1/raw/3e78d6cfb3542dc520570d07648721924cca8b3d/US%2520Zip%2520Codes%2520from%25202016%2520Government%2520Data>

The first one is the postal codes along with the neighborhoods of the Los Angeles.

And the other one is the git containing all the postal codes of USA and their coordinates.

Combining these two data sets gave us the required data set for our analysis.

We use the **BeautifulSoup** package to scrape the required data from websites.

Our required data would have following columns:

PostalCode	Borough	Latitude	Longitude
------------	---------	----------	-----------

We would then use the FourSquare API to get the various famous venues of each area and cluster them using the K-means clustering algorithm.